THE EMBO JOURNAL

Manuscript EMBO-2012-81000

3´-Processing and strand transfer catalyzed by retroviral integrase *in crystallo*

Stephen Hare, Goedele N. Maertens, Peter Cherepanov

Corresponding author: Peter Cherepanov, Cancer Research UK

Review timeline:

Submission date:	05 February 2012
Editorial Decision:	12 March 2012
Revision received:	30 March 2012
Accepted:	02 April 2012

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 12 March 2012

Thank you for submitting your manuscript for consideration by The EMBO Journal. We have now received the reports of three expert referees, which are copied below. I am pleased to inform you that all referees consider the study important and in principle suitable for publication in The EMBO Journal, pending clarification of a number of specific points and satisfactory revision of various aspects related to presentation. With regard to the latter, I would strongly encourage you keep also the broader non-specialist readership of a journal such as ours in mind when editing and proofreading the manuscript, trying to avoid excessive technical jargonry and abbreviations, and where necessary making use of the full-length format of EMBO Journal articles to also include additional explanatory schemes and figures into the main manuscript, as suggested by referee 3. Furthermore, I am not sure that the word 'mechanics' in the title represents an appropriate description of the presented results, and I would appreciate if you could find a more explicit and more generally appealing alternative title for the study. Finally, please make sure to include PDB accession codes into the revised manuscript at this stage.

When preparing your letter of response, please be reminded that our policy to allow only a single round of major revision will necessitate diligent and comprehensive answering, and also bear in mind that this letter will form part of the Peer Review Process File available online to our readers in the case of publication (for more details on our Transparent Editorial Process initiative, please visit our website: http://www.nature.com/emboj/about/process.html).

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor
The EMBO Journal

———

REFEREE REPORTS:

Referee #1 (Remarks to the Author):

This manuscript identifies the crystal structure of several key intermediates of 3' processing and strand-transfer reactions catalyzed by the Prototype Foamy Virus (PFV) Integrase protein. The resolution of these structures is at the level of defining the distinct occupancy of the two divalent metals, as well as the water molecules participating in the nucleophilic attack. Thus the study is quite significant to the field of viral integration, applicable to related fields of nucleotidyl transfers, and has direct implications in the mechanisms of antiviral therapies, specifically targeting the HIV integration process.

Minor comments are listed below:

1) Page 3, line 7. Change "phosphodiesters in host cell DNA" to phosphodiester bonds in the host cell DNA.

2) Page 3, paragraph 2, line 10. "Biomolecular nucleophillic" or bimolecular nucleophilic?

3) Figure 1D. Why is integration driven to a single site within the target DNA? Is this unique to the binding of the target DNA within the crystal?

4) Page 6, second paragraph lines 10-11. The authors state that approximately 50% of the reaction is complete at 120 minutes however there is substantial cleaved product detected at 45 minutes itself. Is this due to variation in the labeling of the substrates/products or an actual difference in the yield of products between T=45 and T=120".

5) Figure 1D. What is the basis for the differential labeling of the N and the R strands? The conditions for labeling with the Clp1 kinase are suboptimal from the published conditions (Jain and Shuman, 2009) for the enzyme (pH 9 vs pH 7.5 or 6.5/ 80{degree sign} C is not optimal for DNA labeling). If there is a justification for the use of these altered conditions, please include it in the manuscript.

Referee #2 (Remarks to the Author):

"Mechanics..." by Hare et al. presents a crystallographic study of the retroviral integrase terminal cleavage reaction. Previously these authors generated the first structures of a retroviral integrase enzyme bound to DNA, a major breakthrough in the field. Here the authors present a series of crystal structures clarifying a previously inaccessible part of the integration process, the 3' processing reaction. Among many novel and interesting points, they show that the coordination sphere around one of the two active site metals is suboptimal, likely contributing to destabilizing the scissile bond in the cleavage reaction. This important paper will be of interest to scientists interested in virology, structural biology, and recombination.

Minor comment.

There was a proliferation of abbreviations in the introduction and throughout (SSC, CCD, TCC, STC, CDC, PFV, INSTI...). This may be unavoidable, but it would be easier on a reader of these could be written out where possible.

Referee #3 (Remarks to the Author):

This paper by Hare, Maertens, and Cherepanov describes a series of crystal structures of the PFV retroviral integrase bound to its DNA substrates and metal ion cofactors at various stages along the integration pathway. One structure of IN bound to intact viral ends is the first time a view prior to end processing has been seen for any retroviral integrase or related DNA transposase. Addition of the metal ions to pre-formed crystals allowed the authors to visualize the reaction as it proceeded, with sufficient resolution to see the movement of metal ions and water molecules in the active site. These structures are important additions to the series of reports already published describing previous steps, and should be of general interest. The correlation between the positions of water molecules in the active site and atoms in INSTIs is a lovely result! Furthermore, the structures clearly explain why the current clinically relevant inhibitors of integrase act at the strand transfer stage of the reaction but not at 3' processing.

That said, one of the main drawbacks of the manuscript as written is its reliance on the reader's previous familiarity with PFV structural work. The manuscript focuses almost immediately on the active site (and in fact, only one strand of the DNA), so it is hard to put the result into context if one isn't already familiar with the context. One way this might be helped would be to move Suppl. Figure 1A into the main body of the text and to provide a simple schematic of the reaction pathway so those not on intimate terms with these can keep track of SSCs, CDCs, TCCs, and STCs. It would also help if the authors were very clear throughout about which of these structures had previously been determined with metal ions - and whether Mg or Mn. (For example, p.5 last paragraph - hasn't the structure of "the TCC poised for integration" been reported already in Maertens et al 2010? Also "the STC resulting from strand transfer in crystallo" in the same paper? What specifically is different here?)

Minor comments
Page 4, middle of page: What is meant by "minimal functional" CDCs, TCCs, and STCs?

Page 8, line 3: It would be helpful to have a brief description of the overall SSC complex, as the fraying of the ends - and the structural reason for this - is not discussed (and in Fig1A, the strands are drawn as unpaired without comment).

Page 8, line 6: If "three bases are unpaired at the non-processed viral DNA end", aren't they unpaired at the processed end as well?

Page 8, 5 lines from bottom: Are all the atoms of the phosphodiester 1.6A further away from the active site, or is this the P-P distance shift?

Page 13, top: It is not clear what is meant by "relatively" loosely bound water and that the enzyme "has a better handle" on the reactants. Could these fuzzy descriptions be replaced with more concrete explanations?

Figure 3A: Please label the bases of the viral DNA.

Figure 4 and page 15, middle of page: It would be easier to follow the discussion of substrate mimicry if the chemical structure of raltegravir were shown somewhere, to complement the 3D view in the figures. Could the F be labelled or a more distinctive colour in the left side images?

Taken together, with the above concerns addressed, the work represents a significant extension of our understanding of the mechanism of retroviral integration, and is likely to be of substantial interest to the wide readership of the EMBO Journal.

1st Revision - authors' response 30 March 2012

We thank all three referees for their positive reports and constructive criticism, which helped us to improve the manuscript.

Our responses to referees' comments are inserted below.

Referee #1:

This manuscript identifies the crystal structure of several key intermediates of 3' processing and strand-transfer reactions catalyzed by the Prototype Foamy Virus (PFV) Integrase protein. The resolution of these structures is at the level of defining the distinct occupancy of the two divalent metals, as well as the water molecules participating in the nucleophilic attack. Thus the study is quite significant to the field of viral integration, applicable to related fields of nucleotidyl transfers, and has direct implications in the mechanisms of antiviral therapies, specifically targeting the HIV integration process.

1) Page 3, line 7. Change "phosphodiesters in host cell DNA" to phosphodiester bonds in the host cell DNA.

Done.

2) Page 3, paragraph 2, line 10. "Biomolecular nucleophillic" or bimolecular nucleophilic?

Thank you for spotting these typos!

3) Figure 1D. Why is integration driven to a single site within the target DNA? Is this unique to the binding of the target DNA within the crystal?

This is an interesting point! Integration *in crystallo* is indeed site-specific (as evident from the analysis of strand transfer products in Figure 1C). But although the target DNA duplex was carefully designed with the actual PFV IN integration site preferences in mind, the selective integration is not due to sequence-specific binding. All sorts of complexes form, but only the symmetric TCC can pack into the crystal lattice, which leads to "site-specific integration" when crystals are soaked in Mn^{2+} or Mg^{2+}. For example, under conditions allowing strand transfer during crystallization (in the presence of Mg^{2+}) we observed selective crystallization of the symmetric STC (Maertens et al., 2010). However, when whole drops were examined, the symmetric product was not even the predominant species (see Supplementary Figure 1 in Maertens et al., 2010). In essence, we purified the symmetric TCC and STC complexes by crystallization. The following was added to the legend of Figure 2: Note that although intasome does not bind target DNA in a sequence-specific fashion, only the symmetric complex crystallizes under the conditions employed (Maertens et al., 2010).

4) Page 6, second paragraph lines 10-11. The authors state that approximately 50% of the reaction is complete at 120 minutes however there is substantial cleaved product detected at 45 minutes itself. Is this due to variation in the labeling of the

substrates/products or an actual difference in the yield of products between T=45 and T=120".

Each lane on these gels contained an individual crystal; the crystals were variable in size, and so was the total amount of DNA in each lane. We consider the ratio between **R** (cleaved) and **r** (uncleaved) bands as an indicator of the reaction progress. The legend to Figure 1B now states: "…based on relative intensities of the R and r bands, the reaction was ~10%, 20% and 50% complete after 20, 45 and 120 min, respectively." We also added a similar note to the panel C: "…based on relative intensities of the T and t bands, the reaction was ~30%, 70%, 90% complete after 30, 120 and 300 sec, respectively." Also, in the results section: "…and the reaction was 50% complete after 120 min (based on relative intensities of R and r bands; Figure 1B)."

5) Figure 1D. What is the basis for the differential labeling of the N and the R strands? The conditions for labeling with the Clp1 kinase are suboptimal from the published conditions (Jain and Shuman, 2009) for the enzyme (pH 9 vs pH 7.5 or 6.5/ 80{degree sign} C is not optimal for DNA labeling). If there is a justification for the use of these altered conditions, please include it in the manuscript.

In the N:R duplex, N displays a single-stranded 5′ end, and is labeled much more efficiently than double-stranded 5′ end of the R strand. Clp1 preformed considerably better than T4 PNK at labeling non-N DNA species, but the problem persisted, even though the thermostable enzyme allowed incubations at 80 ^{o}C. Perhaps some clever "hot-start" protocol could have helped to alleviate this problem; but we have not experimented with this further. The pH reported in the original manuscript was prior to extractions of the samples with phenol. The actual pH of the samples during 5′-end labeling was closer to 7.5-8.0/25^{o}C. Although we have not attempted to measure their true pH at 80^{o}C, we would not expect it to be very far from the optimal pH6.5-7.5. The revised methods section states: "… The protease and SDS were removed by extraction with two changes of phenol-chloroform-isoamyl alcohol (25:24:1, pH8.0). Following two additional extractions with chloroform, the samples were desiccated to dryness and re-dissolved in 10 μl of 10 mM $MgCl_2$, 50 mM Tris-HCl, pH7.4. The DNA products were 5′-end labeled using γ-^{32}P-ATP and thermostable polynucleotide kinase Clp1 at 80 °C (Jain & Shuman, 2009). Labeled DNA species, separated on denaturing urea 17% polyacrylamide gels, were detected by phosphorimaging. RNA-specific 5′-OH kinases such as Clp1 or T4 polynucleotide kinase act on unpaired 5′ ends (Wang et al, 2002). Although both kinases were biased towards phosphorylating the nonreactive strand of the viral DNA duplex (strand N, Figure 1A), which displays a protruding 5′ end following 3′-processing, the thermostable enzyme performed better at labeling other DNA species present in the crystals."

Referee #2:

"Mechanics..." by Hare et al. presents a crystallographic study of the retroviral integrase terminal cleavage reaction. Previously these authors generated the first structures of a retroviral integrase enzyme bound to DNA, a major breakthrough in the field. Here the authors present a series of crystal structures clarifying a previously inaccessible part of the integration process, the 3' processing reaction. Among many novel and interesting points, they show that the coordination sphere around one of the two active site metals is suboptimal, likely contributing to destabilizing the scissile bond in the cleavage reaction. This important paper will be of interest to scientists interested in virology, structural biology, and recombination.

Minor comment.

There was a proliferation of abbreviations in the introduction and throughout (SSC, CCD, TCC, STC, CDC, PFV, INSTI...). This may be unavoidable, but it would be easier on a reader of these could be written out where possible.

We agree. The following abbreviations were removed in the revised version: NTD, CCD, CTD and INSTI. We also replaced SSC and CDC with more intuitive "uncleaved intasome" and "cleaved intasome" (UI and CI). A schematic of the reaction (Figure 1A) will help readers to keep track of the few remaining abbreviations.

Referee #3 (Remarks to the Author):

This paper by Hare, Maertens, and Cherepanov describes a series of crystal structures of the PFV retroviral integrase bound to its DNA substrates and metal ion cofactors at various stages along the integration pathway. One structure of IN bound to intact viral ends is the first time a view prior to end processing has been seen for any retroviral integrase or related DNA transposase. Addition of the metal ions to pre-formed crystals allowed the authors to visualize the reaction as it proceeded, with sufficient resolution to see the movement of metal ions and water molecules in the active site. These structures are important additions to the series of reports already published describing previous steps, and should be of general interest. The correlation between the positions of water molecules in the active site and atoms in INSTIs is a lovely result! Furthermore, the structures clearly explain why the current clinically relevant inhibitors of integrase act at the strand transfer stage of the reaction but not at 3' processing.

That said, one of the main drawbacks of the manuscript as written is its reliance on the reader's previous familiarity with PFV structural work. The manuscript focuses almost immediately on the active site (and in fact, only one strand of the DNA), so it is hard to put the result into context if one isn't already familiar with the context. One way this might be helped would be to move Suppl. Figure 1A into the main body of the text and to provide a simple schematic of the reaction pathway so those not on intimate terms with these can keep track of SSCs, CDCs, TCCs, and STCs.

We agree. A schematic has been added to Figure 1, which should make the paper easier to follow. Also, as the referee suggested, we show the overall structure of the uncleaved intasome (Figure 2A). In addition, the revised manuscript states: "3'-Processing does not grossly affect the overall structure of the intasome, and the r.m.s. deviation between IN main chain atom positions in UI_{Mn} (Figure 2A) and CI_{Mn} (Hare et al., 2010) structures is only ~0.35 Å. While sharing all of the major architectural features described for the cleaved complex (Hare et al., 2010a), the uncleaved intasome structures reveal the IN active site committed for 3′-processing and the configuration of the reaction substrate."

It would also help if the authors were very clear throughout about which of these structures had previously been determined with metal ions - and whether Mg or Mn. (For example, p.5 last paragraph - hasn't the structure of "the TCC poised for integration" been reported already in Maertens et al 2010? Also "the STC resulting from strand transfer in crystallo" in the same paper? What specifically is different here?)

We apologize for less than clear references to our earlier work. To avoid any confusion, the last paragraph of the introduction has been revised to: "Our prior work focused on post-catalytic (CI and STC) or inactivated (TCC lacking 3′-hydroxyls or catalytic metal ions) IN-DNA complexes (Hare et al, 2010a; Maertens et al, 2010). Herein, we present crystal structures of the functional UI and TCC in their ground states committed for 3′-processing and strand transfer, respectively. The structures provide unprecedented insight into the positions of the metal ions and chemical reacting groups in the IN active site, highlight a substrate mimicry utilized by strand transfer inhibitors in their mode of binding to IN and explain why these small molecules are ineffective against the 3′-processing reaction and why inhibitors of 3′-processing have been more difficult to develop." We opted to remove "the STC resulting from strand transfer *in crystallo*" from this paragraph. The complex we referred to originally was STC* (*i.e.* the immediate post-catalytic complex, with two metal ions in the active site), which is different from the STC we published earlier. The difference between STC* and STC is described in the results section.

Minor comments

Page 4, middle of page: What is meant by "minimal functional" CDCs, TCCs, and STCs?

The analogous complexes formed *in vivo* are likely substantially larger and contain various host cell derived components. However, a discussion of this point would defocus the introduction. We opted to delete "minimal functional" from this sentence.

Page 8, line 3: It would be helpful to have a brief description of the overall SSC complex, as the fraying of the ends - and the structural reason for this - is not discussed (and in Fig1A, the strands are drawn as unpaired without comment).

We agree; a new illustration (Figure 2B) makes the structural basis for fraying very clear. In addition, the revised manuscript states: "Three bases are unpaired at the viral DNA end due to the insertion of the IN 3_{10} helix h2 and h2/b5 loop between the reactive and nonreactive viral DNA strands (Figure 2B). The fraying of viral DNA ends prior to 3′-processing is fully consistent with biochemical observations (Katz et al, 2011; Scottoline et al, 1997)."

Page 8, line 6: If "three bases are unpaired at the non-processed viral DNA end", aren't they unpaired at the processed end as well?

The processed end lacks two nucleotides in the reactive strand; and the remaining base (A:T) is unpaired. We opted to delete "non-processed" from this sentence.

Page 8, 5 lines from bottom: Are all the atoms of the phosphodiester 1.6A further away from the active site, or is this the P-P distance shift?

It is correct: the shift is for the P atom. The revised manuscript states: "Two Mn^{2+} ions bind at the active site, inducing relocation of the scissile phosphodiester towards the active site (1.6-Å shift in the position of the P atom; Figure 3B)."

Page 13, top: It is not clear what is meant by "relatively" loosely bound water and that the enzyme "has a better handle" on the reactants. Could these fuzzy descriptions be replaced with more concrete explanations?

The discussion of this point was rather speculative; we opted to delete rather than to expand it in the revised manuscript.

Figure 3A: Please label the bases of the viral DNA.

Done (Now Figure 4A).

Figure 4 and page 15, middle of page: It would be easier to follow the discussion of substrate mimicry if the chemical structure of raltegravir were shown somewhere, to complement the 3D view in the figures. Could the F be labelled or a more distinctive colour in the left side images?

We added new Supplementary Figure 4, which shows the chemical structure of raltegravir. The fluorine atom of the compound is now shown in a darker gray color (see Figure 5 of the revised manuscript).

Taken together, with the above concerns addressed, the work represents a significant extension of our understanding of the mechanism of retroviral integration, and is likely to be of substantial interest to the wide readership of the EMBO Journal.

Acceptance letter 30 March 2012

Thank you for submitting your revised manuscript for our consideration. I have now had a chance to look through it and to assess your responses to the comments raised by the original reviewers, and I am happy to inform you that there are no further objections towards publication in The EMBO Journal.

You shall receive a formal letter of acceptance shortly.

Yours sincerely,
Editor
The EMBO Journal

